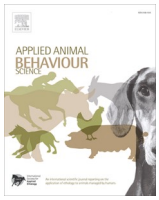
Happiness, Welfare, and Personality in Rhesus Macaques (Macaca mulatta)

Applied Animal Behaviour Science 236 (2021) 105268

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim



Happiness, welfare, and personality in rhesus macaques (*Macaca mulatta*)

Lauren M. Robinson [a,b,*], Natalie K. Waran [c], Ian Handel [d], Matthew C. Leach [e]

[a] Domestication Lab, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine Vienna, Savoyenstraße 1a, A-1160 Vienna, Austria
[b] Language Research Center, Georgia State University, Atlanta, GA, 30034, United States
[c] Faculty of Education, Humanities and Health Science, Eastern Institute of Technology, Hawke's Bay, New Zealand
[d] The Roslin Institute and The Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, EH25 9RG, UK
[e] School of Natural and Environmental Sciences, Agriculture Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

ARTICLE INFO

Keywords:
Coping
Monkey
Psychometrics
Temperament
Validation

ABSTRACT

When we treat an animal's welfare as an individual experience, we should consider the possibility that it may be associated with individual differences in personality. We tested for such associations in 44 socially housed rhesus macaques (*Macaca mulatta*) that lived at the California National Primate Research Center. For each macaque, we obtained ratings on a 16-item welfare questionnaire, a 4-item subjective well-being questionnaire, and a 54-item personality questionnaire, the last of which we used to define each macaque's standing on six personality domains—Confidence, Openness, Dominance, Friendliness, Activity, and Anxiety—identified in an earlier study. Finally, we used focal animal sampling to obtain measures of behavior. We found evidence for interrater agreement for all the welfare items, all but one item from the subjective well-being questionnaire, and all but four items from the personality questionnaire. Using principal components analysis, we found the welfare and subjective well-being survey items loaded together onto a single component. Macaques higher on this dimension received less aggression, engaged in fewer displacement activities (e.g., scratching), and were rated as higher in Confidence, Openness, Dominance, and Friendliness. These results are consistent with reports on chimpanzees and brown capuchin monkeys and constitute further evidence that observer ratings are based on observed behavioral states, suggesting them to be a psychometrically valid way to assess primate welfare.

1. Introduction

In humans, the positive associations between subjective well-being and personality traits related to sociability (e.g., Extraversion) and negative associations between subjective well-being and traits related to anxiety (e.g., Neuroticism) are well-established (DeNeve and Cooper, 1998; Steel et al., 2008). To study this in nonhuman primates, King and Landau (2003) devised the subjective well-being questionnaire, which they modeled on human subjective well-being scales (Diener and Emmons, 1984; Diener et al., 1999; Sandvik et al., 1993). Using this questionnaire, nonhuman primate studies also find that higher subjective well-being is associated with higher sociability and lower anxiety (Inoue-Murayama et al., 2018; Weiss et al., 2009, 2006). For example, adult rhesus macaques that were higher in Confidence and Friendliness and lower in Anxiety were rated as higher in subjective well-being (Weiss et al., 2011a), whereas in infant rhesus macaques, it's been found that those higher in Openness and Dominance and lower in Anxiety are rated higher in subjective well-being (Simpson et al., 2019). Furthermore, as in human studies, the work on subjective well-being in nonhuman primates has shown that it is stable over time (King and Landau, 2003; Weiss et al., 2011b), heritable (Adams et al., 2012; Weiss et al., 2002), and, at least in orangutans, associated with longer life (Weiss et al., 2011a).

Observer ratings have only recently started being used to assess animal welfare (as reviewed by Meagher, 2009 and Whitham and Wielebnowski, 2009). As with studies on animal personality ratings (see reviews by Carter et al., 2013; Gosling, 2001; study by Freeman et al., 2013), before welfare ratings can be used practically, they must be proven to be methodologically sound, for example, by showing them to be reliable (e.g., observers agree, ratings are consistent over time) and to converge with other welfare measures, such as established surveys, behavioral indicators, and physiological responses. Welfare surveys have the potential to be an affordable and fast addition to animal welfare methodologies (Whitham and Wielebnowski, 2009) as animal facilities have limited resources (e.g., staff time) to perform traditional forms of welfare assessment (e.g., formal behavioral observations, hormonal

* Corresponding author at: Savoyenstraße 1a, A-1160 Vienna, Austria.
E-mail address: lauren.robinson@vetmeduni.ac.at (L.M. Robinson).

https://doi.org/10.1016/j.applanim.2021.105268
Received 13 October 2020; Received in revised form 29 January 2021; Accepted 12 February 2021
Available online 17 February 2021

analyses).

Studies thus far have found that observers unfamiliar with individual animals tend to agree on their ratings of farm animal emotion (in sheep: Phythian et al., 2013; in dairy cows: Rousing and Wemelsfelder, 2006; in pigs: Wemelsfelder et al., 2000) and pain-related postures (in dairy cattle: O'Callaghan et al., 2003). Two studies on brown capuchin monkeys *Sapajus apella* (Robinson et al., 2016) and chimpanzees *Pan troglodytes* (Robinson et al., 2017) examined whether observers that were familiar with individual animals could reliably rate their welfare. The researchers found that observers tended to agree on welfare ratings. Moreover, when taken at multiple time points, as they were in a study of chimpanzees, welfare ratings showed evidence of stability along with convergence with behavioral observations (Robinson et al., 2017). Specifically, higher ratings of welfare and subjective well-being were associated with decreased occurrence of regurgitation, coprophagy, and urophagy, which suggests that these ratings of welfare and subjective well-being were based on observed behaviors rather than anthropomorphic projections (Table 4 in Robinson et al., 2017). Additionally, welfare ratings of brown capuchins and chimpanzees were highly correlated with subjective well-being ratings ($rs>0.90$), suggesting that animal welfare and subjective well-being surveys are measures of the same construct (Robinson et al., 2017, 2016). In these studies, it was also found that chimpanzees that were higher in Extraversion and lower in Neuroticism and brown capuchins that were higher in Sociability and Assertiveness and lower in Neuroticism and Attentiveness were judged to have higher welfare.

For the present study, we addressed similar questions about welfare in rhesus macaques *Macaca mulatta* housed at the California National Primate Research Center (CNPRC). Large numbers of rhesus macaques are kept in captivity (Carlsson et al., 2004), and may show signs of distress, such as through the performance of self-injurious behaviors (Lutz et al., 2003). Additionally, ensuring the welfare of animals used in research is also increasingly being considered critical for ensuring the validity of the data collected from such animals (e.g., Richter et al., 2009), as poor welfare can be associated with patho-behavioural and patho-physiological changes in animals and such responses not only compromise welfare but potentially increase the variability, and so decrease the validity, of the data. Given all this, concern over rhesus macaque welfare is considerable and interest in improving their welfare, and that of other research animals, with the goal of producing higher quality science is increasing (Poole, 1997; Prescott and Lidster, 2017). Additionally, the structure of rhesus macaque personality has been previously studied with Weiss et al. (2011b) finding their personality to be made of up of six personality dimensions: Confidence, Openness, Dominance, Friendliness, Activity, and Anxiety. Weiss et al. found these components to be consistent over time and observer and to show excellent internal consistency, making this structure appropriate for use in other studies. Taken together, this made rhesus macaques an ideal species to extend our work on primate personality and welfare.

Our aims were twofold. First, we aimed to test if observers agreed on their ratings of welfare and subjective well-being and if these ratings were associated with another measure of welfare, in this case observed behavior. If this was found to be the case, then we aimed to test for associations between observer ratings of welfare, subjective well-being, and personality, measured using the six component structure from Weiss et al. (2011a).

2. Methods

2.1. Ethical note

This study was non-invasive and complied with the US Animal Welfare Act (2013) and was approved by the University of Edinburgh's Biological Services Unit, A3433-01.

2.2. Subjects

Subjects were 44 group-living rhesus macaques living at the CNPRC in Davis, California, USA. Thirteen of these macaques were male, and the sample ranged in age from 0.92 to 20.97 years (mean ± SD = 8.06 ± 4.88 years). See Supplementary Table 1 for the demographic breakdown of each group. Six macaques were removed from their social group during the study for veterinary reasons. Of these, five macaques were returned to their group during the study period and one was permanently removed.

The macaques lived in one of three groups (N = 16, 15, 13 respectively). Each group was housed in one of three corn cribs, each containing a rectangular cage (2.9 × 2.5 m) with a cylindrical cage (~4 m in diameter) on either side. Each corn crib had a gravel substrate, a metal ceiling, and contained enrichment items, such as plastic barrels and balls. Twice a day the macaques were fed monkey chow (Purina 5045 and 5047) and once a day they were fed additional food items, such as oranges or unshelled peanuts; water was constantly available.

2.3. Instruments

We assessed welfare, subjective well-being, and personality using three questionnaires. The first questionnaire, concerning welfare, has been used to assess welfare in social primate species (Robinson et al., 2017, 2016; Appendix 1)[1] . In the first section of this questionnaire, observers were asked about their professional background and to define the indicators that they use to decide whether an animal had positive and negative welfare. The second section of this questionnaire consisted of 12 items based on five factors identified as being related to animal quality of life (McMillan, 2005), including social relationships, physical health, stress and coping, mental stimulation, and control of the physical and social environment. For this study we also included four new items to assess factors that may affect captive primate welfare, including hair loss (Honess et al., 2005; Kroeker et al., 2014), obesity (Bauer et al., 2011), dominance status (Beisner and Isbell, 2011), and interactions with humans (Davey, 2007; Stoinski et al., 2012). The third section of this questionnaire was used to record demographic information about the animal, including its age and sex. Responses to each item were made on a five-point scale indicating very bad (1) to very good (5) welfare (the wording of the anchors was consistent with the content of the items, e.g., potential responses regarding the frequency of stress ranged from 'never' to 'constantly').

The second questionnaire, concerning subjective well-being, consisted of four items and has been used to assess subjective well-being in nonhuman primates[2] (King and Landau, 2003; Weiss et al., 2006). It asked observers to indicate how happy the animal is, how much pleasure the animal derives from social interactions, how successful the animal is at achieving its goals, and how happy the rater would be if they were that animal for a week. Observers made their responses on a seven-point scale ranging from "Displays either total absence or negligible amounts of the trait or state" to "Displays extremely large amounts of the trait".

The third questionnaire was used to assess the personalities of the subjects. For our purposes, we used the Hominoid Personality Questionnaire[3] (HPQ), which consists of 54 items, each made up of an adjective and up to three descriptive sentences (Weiss et al., 2017, 2009). For example, the item *friendly* is "**FRIENDLY:** Subject often seeks out contact with other monkeys for amiable, genial activities. Subject infrequently initiates hostile behaviors towards other monkeys."

[1] The 16-item animal welfare survey can be found at https://www.drlaurenrobinson.com/surveysdesigned/

[2] The SWB questionnaire can be found at http://extras.springer.com/2011/978−1-4614−0175-9 as weiss_monkey_wellbeing.pdf

[3] The HPQ can be found at http://extras.springer.com/2011/978−1-4614−0175-9/weiss_monkey_personality.pdf

(boldfaced text in the original). Observers made their responses using a seven-point scale that ranges from "Displays either total absence or negligible amounts of the trait" to "Displays extremely large amounts of the trait".

2.4. Ratings

Table 1 summarizes the instruments, observers, and data collected. Welfare and subjective well-being ratings were made between February and June of 2016 by four researchers employed by the CNPRC with between 3 and 16 years of experience using behavior coding and rating scales in primate biobehavioral research. The welfare and subjective well-being questionnaires were completed by one researcher who was acquainted with the subjects (i.e., was able to identify individuals by name prior to the beginning of the study) and three researchers who were unacquainted with the subjects. The researcher who was acquainted with the subjects completed the welfare and subjective well-being questionnaires based on their most recent experience with the macaques. The three researchers who were unacquainted with the subjects were instructed to observe at all individuals within a studied group for a minimum of 30 min a day for five days, resulting in at least 2.5 h of observation time for one entire group, before they completed the questionnaires. To be considerate of their busy schedule, we asked staff to observe at least one group and more only if their work schedules permitted. At the end of the observation period, all four researchers independently completed the welfare and subjective well-being questionnaires. Due to a clerical error, the subjective well-being survey was sent to two of the raters at a later date, which resulted in some animals being rated by fewer observers due to limited time on the part of the observers. Additionally, due to their limited time, we did not ask raters to keep track of how long they spent with each group, meaning we could not account for this variation. Finally, if an animal had been removed from the group and not returned during the study period, then the acquainted researcher was asked to complete ratings based on their most recent experience with the animal; one animal was rated in this manner.

During the study period, again to be considerate of staff work schedules, personality ratings were made by the first author (LMR) and the same researcher that was acquainted with the macaques and filled out the welfare and subjective well-being questions whose ratings were based on their previous experience with the macaques. LMR rated all the macaques on personality after performing the behavioral observations, but before reviewing the other ratings, the behavioral data, and veterinary records and photos of the macaques, the latter two of which were used in two other studies (Altschul et al., 2019; Robinson et al., 2018). We replaced missing values with the mean value for the item (Downey and King, 1998).

2.5. Behavioral observations

All behavioral observations were performed by LMR, a researcher who, at the time of the study, had seven years of experience performing formal animal observations of animal welfare, including research on *Macaca mulatta* and *Macaca fuscata*, at three zoos and the Oregon National Primate Research Center. Prior to observations of the three groups, LMR, who had not worked with the studied individuals before this, spent three days learning to identify individuals within one group. LMR then observed this group for 15 days and observed each macaque once a day, if present, for 15 min. Observation order was randomized on the first day and then rotated so that the first macaque observed on Day 1 was the last macaque observed on Day 2, and so forth. LMR repeated this process until all three groups had been observed.

Observations took place in January to March of 2016. Behavioral observations were conducted using focal animal sampling (Altmann, 1974). Frequencies and durations of behaviors were recorded using Noldus Pocket Observer 3.2 on an Android tablet. We used the ethogram from our previous study of rhesus macaque personality and health (Robinson et al., 2018; Supplementary Table 2). The behaviors recorded included, among others, independent play (duration per observation), locomotor stereotypy (duration per observation), scratch (frequency per observation), and give aggression (frequency per observation). If an individual had to be removed from their social group for any reason, then these animals were observed up until their removal and observations were resumed if they were added back into their social group during the observation period. Factoring this in, each macaque was observed for a mean of 186.85 min ($\pm$ SD = 48.87).

2.6. Statistical analysis

We performed our analyses using version 4.0.0 of R (R Core Team, 2020). To assess the interrater reliabilities of ratings for individual items, we computed two intraclass correlation coefficients for individuals rated by at least two observers: *ICC*(3,1), which estimates the reliability of single ratings, and *ICC*(3,*k*), which assesses the reliability of mean ratings based on *k* raters (Shrout and Fleiss, 1979). The *ICC*s for the welfare items were based on 42 macaques who had been rated by 4 raters ($k = 2.26$). For subjective well-being, *ICC*s were based on ratings of 27 macaques by 2 raters ($k = 2.00$). For personality, *ICC*s were based on ratings of 44 macaques by 2 raters ($k = 2.00$). Items with reliabilities that were equal to or less than zero were omitted from further consideration. Data from all 44 rhesus macaques was used for all the statistics that followed.

We used principal-components analyses (PCA) to reduce the dimensionality of our welfare and subjective well-being measures and to test whether they assessed the same construct. To do so we took means of the welfare and subjective well-being item scores across raters to obtain a single score on each item for each macaque. Using these scores, we conducted an analysis of welfare items and the subjective well-being items. We extracted the number of dimensions recommended by a parallel analysis, which we conducted using the paran function in R (Dinno and Dinno, 2010; Horn, 1965), and an inspection of the scree plot.

After extracting components, we created unit-weighted component scores by assigning a weight of +1 or -1, depending on sign of the loading, and whether it was the highest absolute loading, to loadings greater than or equal to |0.4|, and assigning a weight of 0 to all other loadings (Gorsuch, 1983). We used the results of our PCA to determine the weights for the welfare and subjective well-being scale and the results of an earlier report of rhesus macaque personality to determine the weights used to create scores that represented the personality domains of Confidence, Openness, Dominance, Friendliness, Activity, and Anxiety (Table 1 in Weiss et al., 2011b).

We grouped behaviors into categories to reduce the number of statistical tests and, hence, the Type I error rate. The category "Grooming" included give grooming and self-groom. "Play" included independent play and social play. "Stereotypies" included pacing, back flipping, rocking, and head toss/twirl. "Displacement" included yawn, shake/

Table 1
Summary of instruments, observers, and data collected.

Instrument	Observers	Mean raters	Surveys collected	Item responses	Missing items
Welfare questionnaire	One acquainted, three unacquainted researchers	2.2	97	1552	7
Subjective well-being questionnaire	One acquainted, one unacquainted researcher	1.61	71	122	0
Hominoid Personality Questionnaire	LMR and acquainted researcher	2	88	2376	0

shiver/twitch, and scratch. "Abnormal behavior" included self-suck/clasp, self-bite/injure, salute/eye-poke, coprophagy, teeth grind, and regurgitation and re-ingestion. Several behaviors, i.e., time in proximity, time alone, receive aggression, give aggression, and environment explore were treated separately, as they were not easily categorized; receive grooming was also treated separately as it was made up of the sum of total received grooming from (potentially) multiple other macaques during the observation.

Next, to examine the associations between behaviors, on the one hand, and welfare and subjective well-being on the other, we fit a series of Generalized Linear Mixed Models and linear models. Conceptually, these models are essentially identical to classical procedures such as (repeated measures) ANOVA or regression. However, they allow for more flexibility regarding the assumptions about the residuals (precisely, the distribution of the response given the model) and also can better cope with unbalanced data (Mundry, 2017, 2019), which is why we used them here. For all models, sex = female was used as the reference level. Specifically, we fit three types of models.

First, for the durational response variables (excepting receive grooming) we used Generalized Linear Mixed Effects models with beta error distribution and logit link function using the glmmTMB package in R (Magnusson et al., 2018). The response variables environment explore, time in proximity, time alone, and three of the behavioral categories (i.e., grooming, play, and stereotypies) were divided by (the maximum time observable (900 s) – time out of sight), revealing the proportion of time the individuals spent with each of the six behaviors. We fit one beta regression model for each of the six response variables. These models estimated to what extent the proportion time spent with the respective behaviors (i.e., the response variables) were affected by the predictors in the model. Fixed effects predictors in the models were age, sex, and the welfare and subjective well-being component scores.

Beta regression was not appropriate for the 'receive grooming' behavior as the total time groomed by (potentially) multiple conspecifics in the same session could exceed the maximum time observable. Therefore, we used a linear mixed-effects model using the function lmer of the R package lme4 (Bates et al., 2015) with 'received grooming' as the dependent variable, which we log transformed to account for its right skewed distribution. Fixed effects in the model were age, sex, and the welfare and subjective well-being component scores. We also included total time out-of-sight during observations as a fixed effect to adjust for the fact that some animals were not always visible to the observer.

Lastly, for behaviors for which we counted the number of occurrences, we used Generalized Linear Mixed Effects Models with a Poisson error distribution and log link function using the function glmer of the R package lme4 (Bates et al., 2015) with received aggression, given aggression, and displacement behaviors as the dependent variables (i.e., one model for each of these response variables). Fixed effects in the models were age, sex, and the welfare and subjective well-being component scores; maximum time observable (900 s) – time out of sight was included as an offset variable (log-transformed). One behavioral category, Abnormal, was dropped before analyses due to there being too few instances of the behaviors (51 occurrences across 5 behaviors, 623 observations, and 44 macaques). In all the above described models, Animal ID was included as a random effect to account for variation between individuals.

Following Gelman and Hill (2006), beta and Poisson models were checked for overdispersion using the 'overdisp.test' command from the 'overdisp_correction.r' source code, provided by a statistician (Roger Mundry). Based on the statistician's recommendation, we corrected beta models with dispersion parameters equal or above 1.2 using the 'overdisp_correction.r' source code. If a Poisson model was found to have a dispersion parameter equal or above 1.2, then we switched the distribution to a negative binomial model and again, checked for overdispersion; if the negative binomial distribution produced a dispersion parameter below 1.2, then this result was reported.

Next, we conducted analyses to examine the associations between the component or components derived from the PCA of the welfare subjective well-being items and the personality domains. We first tested whether any of the personality domains were related to the component score or scores. To do so, we used the corr.test function of the psych package (Revelle, 2011) to conduct Spearman rank correlations; we also tested the correlation between personality dimensions using the same method. We then fit a linear model using the lm function; the component score or scores based on the results of the PCA served as dependent variables. The sex, age, and the six personality dimensions were included as fixed effects. However, due to problems with multicollinearity, we instead ran six different linear models with sex, age, and a single personality dimension as the fixed effects and corrected the p-value using a Bonferroni correction. The linear model with all six components and variable inflation values can be found in Supplementary Table 3.

For all the linear and Generalized Linear Mixed Effects models, we transformed the continuous predictor variables by centering them and dividing them by twice the standard deviation; mean and SD of these variables before transformation can be found in Supplementary Table 4. Doing so aids in the interpretation of results as this makes the effect sizes of these variables comparable to the binary variable, i.e., sex (Gelman, 2008). Finally, confidence intervals were calculated using the 'confint' command from the MASS package (Venables and Ripley, 2002) and the boot.lmer from Roger Mundry.

2.7. Data availability

These data and R code have been made publicly available in the supplementary materials section and online at the following DOI: 10.17605/OSF.IO/4NVYJ.

3. Results

3.1. Interrater reliabilities of items

ICC(3,1) estimates for welfare items ranged from 0.04 to 0.61. The *ICC*(3,*k*) estimates for these items ranged from 0.09 to 0.78. *ICC*(3,1) estimates for the subjective well-being items ranged from -0.03 to 0.54. The *ICC*(3,*k*) estimates for these items ranged from -0.06 to 0.70 (Table 2). The mean across welfare and subjective well-being items was 0.25 and 0.39 for the *ICC*(3,1) and *ICC*(3,*k*), respectively. For the HPQ items, *ICC*(3,1) ranged from -0.18 to 0.90 (mean = 0.44) and *ICC*(3,*k*) ranged from -0.46 to 0.94 (mean = 0.54) (Supplementary Table 5). Given these results, we were led to omit one subjective well-being item (*the amount of time that the animal is happy*) and four HPQ items (*stable, solitary, affectionate,* and *protective*).

3.2. Principal-components analysis

Three components had eigenvalues greater than 1. Parallel analysis indicated that only the first was greater than what one would expect from random data at the 95th percentile. The scree plot of the 16 welfare items and the 3 reliable subjective well-being items also indicated that there was a single component. The three reliable subjective well-being items and all but one welfare item (effect of weight) had salient loadings on this component (Table 3). To be consistent with our previous papers (Robinson et al., 2016, 2017), we named this component welfareSWB.

3.3. Generalized mixed effect linear models

In the beta regression models of the proportion of time individuals spent with different behaviors, we found no significant effects of welfareSWB on any of the proportion of times individuals spent performing the durational behaviors (Table 4; the standard deviations, precision

Table 2
Interrater reliability of welfare and SWB items.

Item	ICC(3,1)	ICC(3,k)
Welfare items		
Benefit of rank*	0.61	0.78
Social control	0.53	0.72
Positive/negative experience	0.33	0.52
Stress frequency	0.32	0.51
Negative welfare	0.30	0.50
Stress coping	0.27	0.46
Psychological stimulation	0.22	0.40
Positive welfare	0.22	0.38
Overgrooming*	0.21	0.38
Effect of experience	0.20	0.36
Physical health	0.18	0.33
Number of relationships	0.15	0.29
Environmental control	0.15	0.28
Quality of relationships	0.12	0.24
Interactions with humans*	0.08	0.17
Effect of weight*	0.04	0.09
Subjective well-being items		
SWB Goal achievement	0.54	0.70
SWB Social satisfaction	0.23	0.38
SWB Happiness as animal	0.23	0.38
SWB Time animal is happy	−0.03	−0.06
Mean reliabilities of all items	0.25	0.39

Note. SWB = subjective well-being. Welfare ratings based on 42 rhesus macaques; SWB ratings based on 27 rhesus macaques. Welfare $k = 2.26$ raters per animal; SWB $k = 2.0$.
* new item.

Table 3
16 welfare items, 4 new items, and subjective well-being items.

Item	Loading	h^2
SWB Social Satisfaction	0.90	0.80
Number of relationships	0.89	0.80
Positive/negative experiences	0.88	0.78
Social control	0.88	0.77
SWB Happiness as animal	0.88	0.77
Positive welfare	0.87	0.76
Stress frequency	−0.85	0.72
SWB Goal achievement	0.85	0.72
Negative welfare	−0.84	0.70
Environmental control	0.81	0.65
Quality of relationships	0.78	0.60
*Benefit of rank	0.76	0.58
Stress coping	0.76	0.57
Effect of experience	0.72	0.51
Overgrooming*	−0.64	0.41
Psychological stimulation	0.57	0.32
Physical health	0.50	0.25
Interactions with humans*	0.44	0.20
Effect of weight*	0.21	0.04

Note. N = 44. SWB = subjective well-being. Proportion of variance accounted for = 58 %. h^2=communalities.
* item is one of four newly designed questions.

parameter values, and variance explained by the random effect of subject ID are available in Supplementary Table 6). In the linear model with the amount of grooming received (marginal $R^2 = 0.04$; intercept: b=-2.94, SE = 0.07, P < 0.001, CI 2.14–2.75), we found that male macaques received less grooming (b=-0.68, SE = 0.30, P = 0.02, 95 % CI -1.26 to -0.11) and that the amount of grooming received increased with age (b = 0.59, SE = 0.28, P = 0.034, 95 % CI 0.06–1.13) but we found no significant effects of welfareSWB (b = 0.47, SE = 0.27, P = 0.086, 95 % CI -0.05 to 0.99) or time out of sight (b=-0.16, SE = 0.02, P = 0.35, 95 % CI -0.05 to 0.02). For the frequency behaviors, we found all three to be overdispersed with a Poisson distribution but switching to a negative binomial distribution solved this problem. In the negative binomial

Table 4
Beta models of durational behavior predicted by welfareSWB, sex, and age with animal identity as a random effect.

Model	b	SE	CI	p
Grooming behaviors				
Intercept	−2.94	−2.94	[-3.07,-2.81]	<0.001
WelfareSWB	−0.04	−0.04	[-0.20,0.12]	0.62
Male	−0.25	−0.25	[-0.43,-0.08]	**0.005**
Age	−0.03	−0.03	[-0.19,0.13]	0.695
Mariginal R^2				
0.01				
Locomotor stereotypies*,†				
Intercept	−5.58	0.14	[-5.70,-5.47]	<0.001
WelfareSWB	−0.31	0.19	[-0.47,-0.15]	0.098
Male	−0.12	0.19	[-0.29,0.04]	0.52
Age	0.11	0.17	[-0.04,0.26]	0.53
Mariginal R^2				
0.04				
Play behaviors*,†				
Intercept	−4.81	0.11	[-4.96,-4.66]	<0.001
WelfareSWB	0.11	0.14	[-0.09,0.32]	0.43
Male	0.01	0.16	[-0.21,0.23]	0.96
Age	−0.58	0.16	[-0.80,-0.37]	**<0.001**
Mariginal R^2				
0.08				
Environment explore*,†				
Intercept	−3.25	0.08	[-3.39,-3.12]	<0.001
WelfareSWB	0.00	0.09	[-0.15,0.16]	0.96
Male	−0.11	0.11	[-0.29,0.06]	0.290
Age	−0.60	0.11	[-0.78,-0.42]	**<0.001**
Mariginal R^2				
0.06				
Time alone				
Intercept	−1.75	0.07	[-1.89,-1.61]	<0.001
WelfareSWB	−0.08	0.11	[-0.29,0.13]	0.47
Male	0.18	0.12	[-0.05,0.41]	0.13
Age	0.35	0.11	[0.13,0.56]	**0.002**
Mariginal R^2				
0.04				
Time in proximity				
Intercept	−1.13	0.08	[–1.28,-0.99]	<0.001
WelfareSWB	−0.06	0.12	[-0.30,0.17]	0.61
Male	0.15	0.13	[-0.11,0.41]	0.26
Age	0.27	0.12	[0.02,0.51]	**0.033**
Mariginal R^2				
0.03				

Note. N = 44. Boldface values were significant at P < 0.05; Marginal $R2$ estimates were close to zero and may be unreliable. Sex: female = 0; male = 1.
* Indicates model was corrected for overdispersion.
† CI were not corrected for overdispersion and therefore likely too narrow.

models with frequency of behaviors as the response, we found that macaques with higher welfareSWB score received considerably less aggression and also performed considerably less displacement behaviors (Table 5; Figs. 1 and 2, respectively). We found no significant effects on the amount of aggression given (Table 5). The Spearman correlations of the behavioral categories can be found in Supplementary Table 7.

3.4. Correlation and linear model of welfareSWB and personality dimensions

Higher welfareSWB was significantly correlated with higher Confidence, Openness, Dominance, and Friendliness (Table 6). Activity and Anxiety were not correlated with WelfareSWB. The Spearman correlations between the personality components can be found in Supplementary Table 8. In the linear model that included Confidence, we found that macaques that were younger and scored higher in Confidence were rated as having higher welfareSWB (Table 7). In the linear model that included Openness, we found that macaques that scored higher in Openness were rated as having higher welfareSWB. In the linear model that included Dominance, we found that macaques that were higher in Dominance were rated as having higher welfareSWB. In the linear model

Table 5
Negative binomial models of frequency behavior predicted by welfareSWB, sex, and age with time out of sight as an offset and animal identity as a random effect.

Model	*b*	SE	CI	p
Receive aggression				
Intercept	−7.46	0.15	[-7.79,-7.18]	<0.001
WelfareSWB	−1.26	0.27	[-1.82,-0.72]	**<0.001***
Male	−0.48	0.29	[-1.08,0.06]	0.095
Age	−1.77	0.31	[-2.46,-1.18]	**<0.001**
Mariginal R^2				
0.33				
Give aggression				
Intercept	−7.46	0.18	[-7.83,-7.15]	<0.001
WelfareSWB	0.49	0.31	[-0.08,1.11]	0.12
Male	−0.25	0.33	[-0.90,0.39]	0.45
Age	0.05	0.31	[-0.55,0.65]	0.88
Mariginal R^2				
0.30				
Displacement behaviors				
Intercept	−5.68	0.07	[-5.82,-5.55]	<0.001
WelfareSWB	−0.35	0.12	[-0.56,-0.13]	**0.003**
Male	0.15	0.13	[-0.11,0.40]	0.25
Age	−0.23	0.12	[-0.48,-0.01]	0.051
Mariginal R^2				
0.06				

Note. N = 44. Boldface values were significant at P < 0.05; Marginal *R*2 estimates were close to 0 and may be unreliable. Sex: female = 0; male = 1.
* value was significant at P < 0.05 after correction.

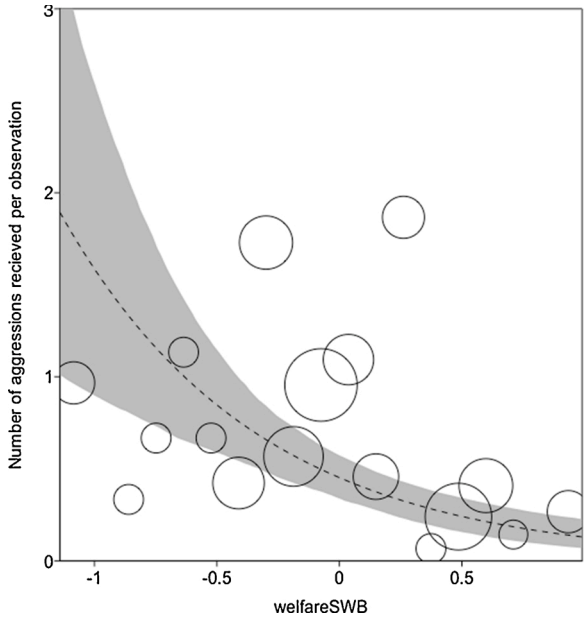


Fig. 1. Number of aggressions received in an observation time of 900 s, which roughly equals the average observation time, as a function of welfareSWB score. Bubbles indicate the average number of aggressions received per binned welfareSWB score whereby the area of the bubbles depicts the number of data points per bin (N = 14 to 89). The dashed line and grey polygon depict the fitted model and its confidence limits (with all other effects centered to a mean of zero). Note that the estimated number of aggressions received decreased from almost 0.97 to ca. 0.27 from individuals lowest to highest on the welfareSWB score.

that included Friendliness, we found that macaques that were higher in Friendliness were rated as having higher welfareSWB. We found no relationship between Activity, Anxiety, or age and sex in the models that included these components. The relationship between personality and welfareSWB is depicted in Fig. 3.

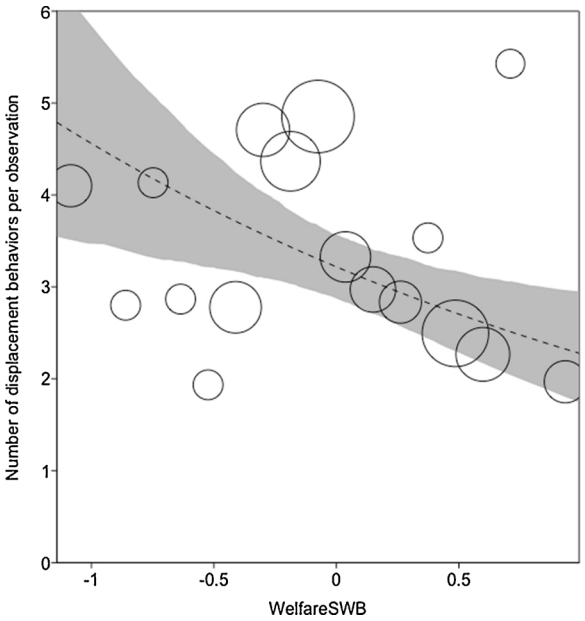


Fig. 2. Number of displacement behaviors performed in an observation time of 900 s, which roughly equals the average observation time, as a function of welfareSWB score. Bubbles indicate the average number of displacement behaviors per binned welfareSWB score whereby the area of the bubbles depicts the number of data points per bin (N = 14 to 89) The dashed line and grey polygon depict the fitted model and its confidence limits (with all other effects centered to a mean of zero). Note that estimated number of displacement behaviors approximately halved from ca 4.1 to ca. 1.96 from individuals lowest to highest on the welfareSWB score.

Table 6
Spearman correlations of rhesus macaque personality components and welfareSWB component.

Component	*r*	95 % CI	P
Confidence	**0.82**	[0.68,0.90]	<0.001
Openness	**0.40**	[0.12,0.63]	0.025
Dominance	**0.75**	[0.59,0.86]	<0.001
Friendliness	**0.56**	[0.31,0.73]	<0.001
Activity	0.31	[0.01,0.55]	0.084
Anxiety	−0.24	[-0.50,0.06]	0.12

Note. N = 44. Boldface values were significant at P < 0.05.

4. Discussion

We found that all but one subjective well-being item and four personality items had positive interrater reliabilities and that welfare and subjective well-being items loaded onto a single component (welfareSWB). Macaques with higher scores on this component performed significantly fewer displacement behaviors and received less aggression. Higher welfareSWB scores were also significantly correlated with higher Confidence, Openness, Dominance, and Friendliness, all of which were significantly associated with welfareSWB in individual linear models.

Previous studies have shown that observers agree on their scoring of behaviors that indicate welfare status, such as pain measured with grimace scales (as reviewed by Descovich, 2017; Leach et al., 2012) and lameness measured with gait scores (Brenninkmeyer et al., 2007). Our findings here show similarly that observers agree on ratings of welfare, which is compelling given we had a mix of acquainted and unacquainted observers. Perhaps, when observers are experienced with the species, as our observers were, they do not have to know the individual for long to be able to agree on their welfare status, though of course increased time with the animals would likely have increased the degree of agreement. That higher welfareSWB was associated with performing fewer

Table 7
Linear models of rhesus macaque welfareSWB predicted by each personality component.

Model	*b*	SE	CI	p
	Confidence model			
Intercept	0.00	1.09	[-2.19,2.19]	1.00
Age	−11.38	1.94	[-15.30,-7.46]	**<0.001**
Male	−0.01	2.06	[-04.16,4.14]	1.00
Confidence	24.94	1.88	[21.14,28.73]	**<0.001**
R^2				
0.83				
	Openness model			
Intercept	−1.57	2.33	[-6.27,3.14]	0.50
Age	5.29	5.59	[-6.00,16.58]	0.35
Male	5.30	4.43	[-3.66,14.26]	0.24
Openness	14.84	5.41	[3.90,25.77]	**0.009**
R^2				
0.21				
	Dominance model			
Intercept	−1.26	1.54	[-4.36,1.85]	0.42
Age	−2.77	2.69	[-8.21,2.66]	0.31
Male	4.25	2.90	[-1.61,10.10]	0.15
Dominance	21.42	2.60	[16.16,26.67]	**<0.001**
R^2				
0.65				
	Friendliness model			
Intercept	−0.59	2.10	[-4.83,3.66]	0.78
Age	−1.17	3.78	[-8.82,6.47]	0.76
Male	1.99	3.96	[-6.02,10.00]	0.62
Friendliness	15.55	3.69	[8.09,23.00]	**<0.001**
R^2				
0.35				
	Activity model			
Intercept	−0.75	2.43	[-5.66,4.16]	0.76
Age	2.33	6.04	[-9.87,14.54]	0.70
Male	2.54	4.58	[-6.73,11.80]	0.58
Activity	10.68	6.00	[-1.45,22.80]	0.083
R^2				
0.13				
	Anxiety model			
Intercept	−0.12	2.37	[-4.92,4.68]	0.96
Age	−12.86	5.11	[-23.19,-2.52]	0.016
Male	0.41	4.56	[-8.81,9.64]	0.93
Anxiety	−12.10	4.95	[-22.10,-2.10]	0.019
R^2				
0.18				

Note. N = 44. Boldface values are significant at P < 0.05 after Bonferroni correction. Sex: female = 0; male = 1.

displacement behaviors and receiving less aggression, which have traditionally been used as negative welfare indicators (Troisi, 2002), suggests that observers may have used these behavioral indicators to make their judgements.

The observers did not appear to use play behavior when making their ratings of welfare and subjective well-being. This may be because play behavior is difficult to judge and/or less commonly seen (Mellor and Beausoleil, 2015; Yeates and Main, 2008), may not be a reliable behavioral indicator of positive welfare (Yamanashi et al., 2018), or that the survey questions may have been inappropriate for capturing indicators of positive welfare, suggesting further modifications and studies are needed. However, it's worth noting that while we found no association between welfareSWB and what are considered to be behavioral indicators of positive welfare, we are not alone in this. Play for example, is often discussed as a positive welfare indicator (Held and Špinka, 2011). Yet, Ahloy-Dallaire et al. (2018) found that reduced welfare can result in reduced play in animals and children, though they note cases where the opposite is true, but could not find enough evidence to determine if it is an indicator of positive welfare or merely neutral welfare. Perhaps the lack of support for these measures is unsurprising given how little research is available on them. A recent review by Lawrence et al. (2019) showed that there are only 71 papers that use the keywords "positive welfare" and "animal" in the title, key words, and abstract. Given the limited literature and mixed findings, it is possible that it is not the survey that is problematic but instead the behaviors used to indicate positive welfare.

Some researchers discuss animal welfare as being made up of different factors (McMillan, 2005) or domains (Mellor and Beausoleil, 2015). However, the view that welfare is multidimensional (Botreau et al., 2007; Fraser, 1995) was not borne out in this study as we found a unidimensional structure of the welfare and subjective well-being items. This unidimensionality of welfare was also found in studies of brown capuchins (Robinson et al., 2016) and chimpanzees (Robinson et al., 2017), i.e., these studies also found a single component that captured welfare and subjective well-being variance. This suggests that at the level of the individual, welfare may not be separated into different dimensions or factors (e.g., social, physical, psychological), but that these different aspects of welfare all contribute to a single welfare dimension.

We found that macaques with higher welfareSWB scores were significantly higher in Confidence, Openness, Dominance, and Friendliness. Similar associations were shown between personality and welfare in brown capuchins: younger animals and those higher in Assertiveness and Sociability, and lower in Attentiveness and Neuroticism, had higher welfareSWB (Robinson et al., 2016). Previous work in semi-free-ranging rhesus macaques (Weiss et al., 2011b) found that, as in this study, higher Confidence and Friendliness were associated with subjective well-being, though they also found higher levels to be related to lower Anxiety. The differences between results may be due to our inclusion of questions on more traditional aspects of welfare, such as physical health and stress, the fact that one sample was derived from a research center and the other sample was derived from a semi-free-ranging population, or that we lacked the statistical power to find these associations.

It's important that we acknowledge the limitations of our study. For example, all observations were performed by LMR and thus we were unable to test interobserver reliability. As a higher number of ratings can be averaged across observers, which results in more reliable data (Vazire et al., 2007), the clerical error that led to 17 macaques only being rated once on the subjective well-being survey was less than ideal. Additionally, the reliability of some of the items were lower than in past studies, likely due to the lower number or raters and/or their limited time with the animals. As such, future research replicating these findings, ideally with a higher number of familiar and unfamiliar observers, will be valuable for testing the generalizability of these results. It is worth noting that in spite of a lower number of ratings, primarily unfamiliar observers, and some animals only being rated once on subjective well-being, we still found the majority of the items to load onto the structure produced by PCA, where only items with reliable variance would load at our criteria of |0.40|, resulting in similar structure to those found in our other studies of brown capuchins and chimpanzees (Robinson et al., 2016, 2017). That this structure was related to observed behaviors suggests that these ratings and items were valuable and appropriate to include in our analyses. An additional limitation was that we had relatively few observations for such a behaviorally complex species. This means that null findings between rating and behavioral data may simply reflect the inability to separate signal from noise—i.e., consistent behavioral differences (which should reflect personality ratings) from fluctuating behavioral expressions (which might reflect situational factors not captured by the ratings). A longer observation period and a larger sample size would be beneficial to future studies of this topic.

The significant relationship between higher Confidence and higher welfareSWB in this sample may have come about because, as we showed in a previous report, rhesus macaques (including the animals from the present sample) that are higher in Confidence tended to be injured less frequently (Robinson et al., 2018). Other studies have shown that personality differences are related to nonhuman primate welfare, whether measured using behavioral indicators (e.g., Gottlieb et al., 2013, 2015; Peterson et al., 2017), ratings of subjective well-being (e.g., King and Landau, 2003), or health (e.g., Jin et al., 2013; Weiss et al., 2013). Taken

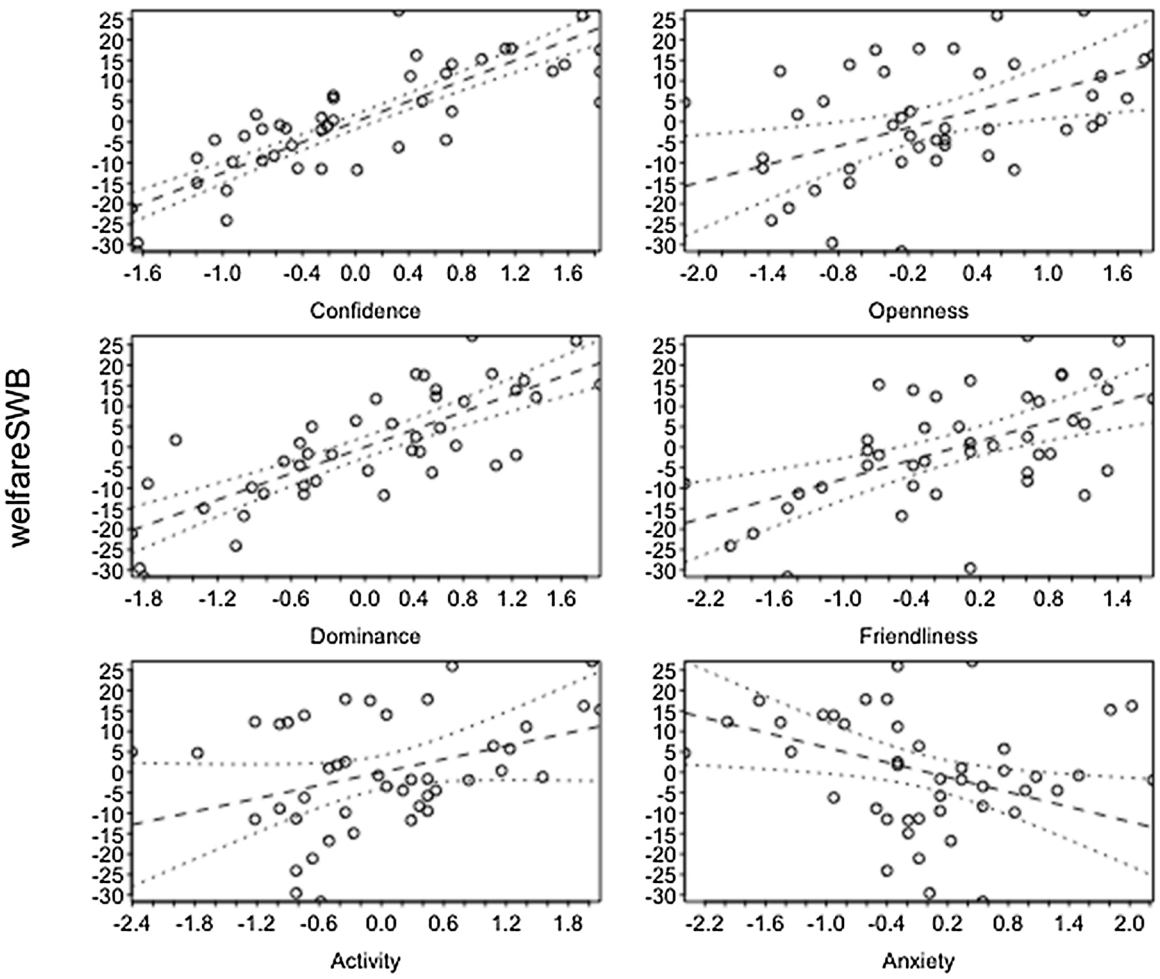


Fig. 3. Relationship between each of the six personality component scores (x-axes) and welfareSWB component score (y-axes). Each dot represents one individual rhesus macaque (N = 44). Dashed and dotted lines depict the fitted model and its confidence limits.

together, such findings suggest that individual differences in personality and welfare reflect a need to individualize the care of captive animals, such as by providing additional veterinary exams and care to those more likely to become injured or receive more aggression.

Though we feel more validation work needs to be done, for example through testing the long-term stability of ratings and how much time observers need know an animal to produce reliable welfare ratings, we see multiple lines of application for both welfare surveys and the relationship between welfare and personality First, the identification of animals with reduced welfare, which can then be used to improve it. While physical injuries, like broken bones and lacerations, are commonly seen and treated quickly, it is be more difficult to identify when other aspects of welfare, such as from reduced ability to cope with their social environment, are compromised. The welfareSWB survey covers multiple aspects of welfare, including psychological welfare, and thus the scores derived from it may be a useful indicator of overall welfare that can be used to find such individuals and then intervene, such as by moving animals to a new social group or more appropriate environment for their needs. This brings us to our second application, to track welfare over time and in response to changes, such as the addition of a new group member. Although the validation methods and statistics within this paper may be complex, the actual entry and tracking of welfare scores over time is far less so. Funded projects, like WelfareTrak (Chicago Zoological Society, 2021) and Animal Welfare Assessment Grid (Justice et al., 2017), show that this can be made user friendly with observers simply answering questions online and the program tracking and graphing these scores, suggesting that our survey could similarly be used. Regarding personality, we previously mentioned that the relationship between personality and welfare could be used to individualize care, such as provision of additional enrichment or more frequent veterinary exams. In line with this, we suggest that studies testing if personality predicts how beneficial different types of enrichment (e.g.., social, physical, nutritional, occupational, and sensory; Bloomsmith et al., 1991) are for animals, may be useful improving their welfare. For example, if less extraverted animals benefit more from physical enrichment than social, then these individuals may be best kept in smaller groups with more items within their enclosure. As welfare is an individual experience, we believe that moving towards individualized care that consider who animals are (i.e., their personality) will be beneficial to their welfare.

Observers that are experienced and knowledgeable about a species, such as research assistants, animal technicians, and zoo keepers, may provide reliable and valid animal welfare scores, even if they spend relatively little time getting to know individual animals. Further, by considering animal welfare as an individual experience that is connected to personality, we begin to see the complex relationships that determine and influence an individual animal's welfare and happiness. This leaves us not only better prepared to study these needs, but to meet them in the future through individualized care.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

We thank the researchers and staff who completed the welfare and personality questionnaires. We also thank Alexander Weiss, Ashleigh Day, Kyle Bone, Laura Del Rosso, and Laura Calonder for feedback and help with this project. Thank you to Roger Mundry for statistics and R support. Thank you, too, to Kirstin McIlvaney at the Scotland's Rural College for her help with Observer. We would like to thank the California National Primate Research Center (NIH grant P51OD011107) for hosting this project.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.applanim.2021.105268.

References

Adams, M.J., King, J.E., Weiss, A., 2012. The majority of genetic variation in orangutan personality and subjective well-being is nonadditive. Behav. Genet. 42 (4), 675–686. https://doi.org/10.1007/s10519-012-9537-y.

Ahloy-Dallaire, J., Espinosa, J., Mason, G., 2018. Play and optimal welfare: does play indicate the presence of positive affective states? Behav. Processes 156, 3–15. https://doi.org/10.1016/j.beproc.2017.11.011.

Altmann, J., 1974. Observational study of behavior: sampling methods. Behaviour 49 (3–4), 227–266. https://doi.org/10.1163/156853974X00534.

Altschul, D.M., Robinson, L.M., Coleman, K., Capitanio, J.P., Wilson, V.A.D., 2019. An exploration of the relationships among facial dimensions, age, sex, dominance status, and personality in Rhesus macaques (Macaca mulatta). Int. J. Primatol. https://doi.org/10.1007/s10764-019-00104-y.

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. https://doi.org/10.18637/jss.v067.i01.

Bauer, S.A., Arndt, T.P., Leslie, K.E., Pearl, D.L., Turner, P.V., 2011. Obesity in rhesus and cynomolgus macaques: a comparative review of the condition and its implications for research. Comp. Med. 61 (6), 514–526. Retrieved from. http://www.ncbi.nlm.nih.gov/pubmed/22330579.

Beisner, B.A., Isbell, L.A., 2011. Factors affecting aggression among females in captive groups of rhesus macaques (*Macaca mulatta*). Am. J. Primatol. 73, 1–8.

Bloomsmith, M., Brent, L., Schapiro, S., 1991. Guidelines for developing and managing an environmental enrichment program for nonhuman primates. Lab. Anim. Sci. 41, 372–377.

Botreau, R., Veissier, I., Butterworth, A., Bracke, M.B.M., Keeling, L.J., 2007. Definition of criteria for overall assessment of animal welfare. Anim. Welf. 16 (2), 225–228.

Brenninkmeyer, C., Dippel, S., March, S., Brinkmann, J., Winckler, C., Knierim, U., 2007. Reliability of a subjective lameness scoring system for dairy cows. Anim. Welf. 16, 127–129.

Carlsson, H., Schapiro, S.J., Farah, I., Hau, J., 2004. Use of primates in research: a global overview. Am. J. Primatol. 63 (4), 225–237. https://doi.org/10.1002/ajp.20054.

Carter, A.J., Feeney, W.E., Marshall, H.H., Cowlishaw, G., Heinsohn, R., 2013. Animal personality: what are behavioural ecologists measuring? Biol. Rev. 88 (2), 465–475. https://doi.org/10.1111/brv.12007.

Chicago Zoological Society. (n.d.). What is Welfare Trak? Retrieved from https://welfaretrak.org/What-is-WelfareTrak.aspx.

Davey, G., 2007. Visitors' effects on the welfare of animals in the zoo: a review. J. Appl. Anim. Welf. Sci. 10 (2), 169–183. https://doi.org/10.1080/10888700701313595.

DeNeve, K.M., Cooper, H., 1998. The happy personality: a meta-analysis of 137 personality traits and subjective well-being. Psychol. Bull. 124 (2), 197–229. https://doi.org/10.1037/0033-2909.124.2.197.

Descovich, K., 2017. Facial expression: an under-utilised tool for the assessment of welfare in mammals. ALTEX. https://doi.org/10.14573/altex.1607161.

Diener, E., Emmons, R.A., 1984. The independence of positive and negative affect. J. Pers. Soc. Psychol. 47 (5), 1105–1117. https://doi.org/10.1037/0022-3514.47.5.1105.

Diener, E., Suh, E.M., Lucas, R.E., Smith, H.L., 1999. Subjective well-being: three decades of progress. Psychol. Bull. 125 (2), 276–302. https://doi.org/10.1037/0033-2909.125.2.276.

Dinno, A., Dinno, M., 2010. R Package "paran.".

Downey, R.G., King, C., 1998. Missing data in Likert ratings: a comparison of replacement methods. J. Gen. Psychol. 125 (2), 175–191. https://doi.org/10.1080/00221309809595542.

Fraser, D., 1995. Science, values and animal welfare: exploring the "inextricable connection". Anim. Welf. 4, 103–117.

Freeman, H.D., Brosnan, S.F., Hopper, L.M., Lambeth, S.P., Schapiro, S.J., Gosling, S.D., 2013. Developing a comprehensive and comparative questionnaire for measuring personality in chimpanzees using a simultaneous top-down/bottom-up design. Am. J. Primatol. 75 (10), 1042–1053. https://doi.org/10.1002/ajp.22168.

Gelman, A., 2008. Scaling regression inputs by dividing by two standard deviations. Stat. Med. 27 (15), 2865–2873. https://doi.org/10.1002/sim.3107.

Gelman, A., Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511790942.

Gorsuch, R.L., 1983. Factor Analysis. Retrieved from, 2nd ed. L. Erlbaum Associates, Hillsdale, NJ. https://market.android.com/details?id=book-Y5kfvgAACAAJ.

Gosling, S.D., 2001. From mice to men: What can we learn about personality from animal research? Psychol. Bull. 127 (1), 45–86. https://doi.org/10.1037/0033-2909.127.1.45.

Gottlieb, D.H., Capitanio, J.P., McCowan, B.J., 2013. Risk factors for stereotypic behavior and self-biting in rhesus macaques (*Macaca mulatta*): Animal's history, current environment, and personality. Am. J. Primatol. 75 (10), 995–1008. https://doi.org/10.1002/ajp.22161.

Gottlieb, D.H., Maier, A., Coleman, K., 2015. Evaluation of environmental and intrinsic factors that contribute to stereotypic behavior in captive rhesus macaques (*Macaca mulatta*). Appl. Anim. Behav. Sci. 171, 184–191. https://doi.org/10.1016/j.applanim.2015.08.005.

Held, S.D.E., Špinka, M., 2011. Animal play and animal welfare. Anim. Behav. https://doi.org/10.1016/j.anbehav.2011.01.007.

Honess, P.E., Gimpel, J., Wolfensohn, S.E., Mason, G.J., 2005. Alopecia scoring: the quantitative assessment of hair loss in captive macaques. Altern. Lab. Anim. 33 (3), 193–206.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185. https://doi.org/10.1007/BF02289447.

Inoue-Murayama, M., Yokoyama, C., Yamanashi, Y., Weiss, A., 2018. Common marmoset (*Callithrix jacchus*) personality, subjective well-being, hair cortisol level and AVPR1a, OPRM1, and DAT genotypes. Sci. Rep. 8 (1), 10255 https://doi.org/10.1038/s41598-018-28112-7.

Jin, J., Su, Y., Tao, Y., Guo, S., Yu, Z., 2013. Personality as a predictor of general health in captive golden snub-nosed monkeys (*Rhinopithecus roxellana*). Am. J. Primatol. 75 (6), 524–533. https://doi.org/10.1002/ajp.22127.

Justice, W.S.M., O'Brien, M.F., Szyszka, O., Shotton, J., Gilmour, J.E.M., Riordan, P., Wolfensohn, S.E., 2017. Adaptation of the animal welfare assessment grid (AWAG) for monitoring animal welfare in zoological collections. Vet. Rec. 181 (6), 143. https://doi.org/10.1136/vr.104309.

King, J.E., Landau, V.I., 2003. Can chimpanzee (Pan troglodytes) happiness be estimated by human raters? J. Res. Pers. 37 (1), 1–15. https://doi.org/10.1016/S0092-6566(02)00527-5.

Kroeker, R., Bellanca, R.U., Lee, G.H., Thom, J.P., Worlein, J.M., 2014. Alopecia in three macaque species housed in a laboratory environment. Am. J. Primatol. 76 (4), 325–334. https://doi.org/10.1002/ajp.22236.

Lawrence, A.B., Vigors, B., Sandøe, P., 2019. What is so positive about positive animal welfare?—a critical review of the literature. Animals. https://doi.org/10.3390/ani9100783.

Leach, M.C., Klaus, K., Miller, A.L., Scotto di Perrotolo, M., Sotocinal, S.G., Flecknell, P. A., 2012. The assessment of post-vasectomy pain in mice using behaviour and the Mouse Grimace Scale. PLoS One 7 (4), e35656. https://doi.org/10.1371/journal.pone.0035656.

Lutz, C.K., Well, A., Novak, M.A., 2003. Stereotypic and self-injurious behavior in rhesus macaques: a survey and retrospective analysis of environment and early experience. Am. J. Primatol. 60 (1), 1–15. https://doi.org/10.1002/ajp.10075.

Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Mächler, M., et al., 2018. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 9 (2), 378–400. https://doi.org/10.3929/ETHZ-B-000240890.

McMillan, F., 2005. Mental wellness: the concept of quality of life in animals. Mental Health and Well-Being in Animals. Blackwell Publishing Ltd.

Meagher, R.K., 2009. Observer ratings: validity and value as a tool for animal welfare research. Appl. Anim. Behav. Sci. 119 (1–2), 1–14. https://doi.org/10.1016/j.applanim.2009.02.026.

Mellor, D., Beausoleil, N., 2015. Extending the "Five Domains" model for animal welfare assessment to incorporate positive welfare states. Anim. Welf. 24 (3), 241–253. https://doi.org/10.7120/09627286.24.3.241.

Mundry, R., 2017. From nonparametric tests to mixed models: a brief overview of statistical tools frequently used in comparative psychology. APA Handbook of Comparative Psychology: Basic Concepts, Methods, Neural Substrate, and Behavior. American Psychological Association, Washington, pp. 157–177. https://doi.org/10.1037/0000011-008.

Mundry, R., 2019. Developments in statistical methods applied over four decades of research in the Taï Chimpanzee Project. The Chimpanzees of the Taï Forest. Cambridge University Press, pp. 28–43. https://doi.org/10.1017/9781108674218.003.

O'Callaghan, K.A., Cripps, P.J., Downham, D.Y., Murray, R.D., 2003. Subjective and objective assessment of pain and discomfort due to lameness in dairy cattle. Anim. Welf. 12 (4), 605–610.

Peterson, E.J., Worlein, J.M., Lee, G.H., Dettmer, A.M., Varner, E.K., Novak, M.A., 2017. Rhesus macaques (*Macaca mulatta*) with self-injurious behavior show less behavioral anxiety during the human intruder test. Am. J. Primatol. 79 (1) https://doi.org/10.1002/ajp.22569 e22569.

Phythian, C., Michalopoulou, E., Duncan, J., Wemelsfelder, F., 2013. Inter-observer reliability of qualitative behavioural assessments of sheep. Appl. Anim. Behav. Sci. 144 (1–2), 73–79. https://doi.org/10.1016/j.applanim.2012.11.011.

Poole, T., 1997. Happy animals make good science. Lab. Anim. 31 (2), 116–124. https://doi.org/10.1258/002367797780600198.

Prescott, M.J., Lidster, K., 2017. Improving quality of science through better animal welfare: the NC3Rs strategy. Lab Anim. 46 (4), 152–156. https://doi.org/10.1038/laban.1217.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria https://doi.org/http://www.R-project.org/.
Revelle, M.W., 2011. Package 'Psych. October, pp. 1–250.
Robinson, L.M., Waran, N.K., Leach, M.C., Morton, F.B., Paukner, A., Lonsdorf, E., et al., 2016. Happiness is positive welfare in brown capuchins (Sapajus apella). Appl. Anim. Behav. Sci. 181, 145–151. https://doi.org/10.1016/j.applanim.2016.05.029.
Robinson, L.M., Altschul, D.M., Wallace, E.K., Úbeda, Y., Llorente, M., Machanda, Z., et al., 2017. Chimpanzees with positive welfare are happier, extraverted, and emotionally stable. Appl. Anim. Behav. Sci. 191, 90–97. https://doi.org/10.1016/j.applanim.2017.02.008.
Robinson, L.M., Coleman, K., Capitanio, J.P., Gottlieb, D.H., Handel, I.G., Adams, M.J., et al., 2018. Rhesus macaque personality, dominance, behavior, and health. Am. J. Primatol. 80 (2), e22739 https://doi.org/10.1002/ajp.22739.
Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. Appl. Anim. Behav. Sci. 101 (1–2), 40–53. https://doi.org/10.1016/j.applanim.2005.12.009.
Sandvik, E., Diener, E., Seidlitz, L., 1993. Subjective well-being - the convergence and stability of self-report and non-self-report measures. J. Pers. 61 (3), 318–342. https://doi.org/10.1111/j.1467-6494.1993.tb00283.x.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86 (2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420.
Simpson, E.A., Robinson, L.M., Paukner, A., 2019. Infant rhesus macaque (*Macaca mulatta*) personality and subjective well-being. PLoS One 14 (12). https://doi.org/10.1371/journal.pone.0226747 e0226747.
Steel, P., Schmidt, J., Shultz, J., 2008. Refining the relationship between personality and subjective well-being. Psychol. Bull. 134 (1), 138–161. https://doi.org/10.1037/0033-2909.134.1.138.
Stoinski, T.S., Jaicks, H.F., Drayton, L.A., 2012. Visitor effects on the behavior of captive Western lowland gorillas: the importance of individual differences in examining welfare. Zoo Biol. 31 (5), 586–599. https://doi.org/10.1002/zoo.20425.
Troisi, A., 2002. Displacement activities as a behavioral measure of stress in nonhuman primates and human subjects. Stress 5 (1), 47–54. https://doi.org/10.1080/10253890290001237 8.
Vazire, S., Gosling, S.D., Dickey, A.S., Schapiro, S.J., 2007. Measuring personality in nonhuman animals. In: Robins, R., Fraley, R.C., Krueger, R.F. (Eds.), Handbook of Research Methods in Personality Psychology. The Guildford Press, New York, pp. 190–207.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics With S, fourth. Springer, New York.
Weiss, A., King, J.E., Enns, R.M., 2002. Subjective well-being is heritable and genetically correlated with dominance in chimpanzees (*Pan troglodytes*). J. Pers. Soc. Psychol. 83 (5), 1141–1149. https://doi.org/10.1037/0022-3514.83.5.1141.
Weiss, A., King, J.E., Perkins, L., 2006. Personality and subjective well-being in orangutans (*Pongo pygmaeus* and *Pongo abelii*). J. Pers. Soc. Psychol. 90 (3), 501–511. https://doi.org/10.1037/0022-3514.90.3.501.
Weiss, A., Inoue-Murayama, M., Hong, K.-W., Inoue, E., Udono, T., Ochiai, T., et al., 2009. Assessing chimpanzee personality and subjective well-being in Japan. Am. J. Primatol. 71 (4), 283–292. https://doi.org/10.1002/ajp.20649.
Weiss, A., Adams, M.J., King, J.E., 2011a. Happy orang-utans live longer lives. Biol. Lett. 7 (6), 872–874. https://doi.org/10.1098/rsbl.2011.0543.
Weiss, A., Adams, M.J., Widdig, A., Gerald, M.S., 2011b. Rhesus macaques (Macaca mulatta) as living fossils of hominoid personality and subjective well-being. J. Comp. Psychol. 125 (1), 72–83. https://doi.org/10.1037/a0021187.
Weiss, A., Gartner, M.C., Gold, K.C., Stoinski, T.S., 2013. Extraversion predicts longer survival in gorillas: an 18-year longitudinal study. Proc. R. Soc. B: Biol. Sci. 280 (1752), 20122231 https://doi.org/10.1098/rspb.2012.2231.
Weiss, A., 2017. Exploring factor space (and other adventures) with the hominoid personality questionnaire. In: Vonk, J., Weiss, A., Kuczaj, S. (Eds.), Personality in Nonhuman Animals. Springer, Cham, pp. 19–38. https://doi.org/10.1007/978-3-319-59300-5_2.
Wemelsfelder, F., Hunter, E., Mendl, M., Lawrence, A., 2000. The spontaneous qualitative assessment of behavioural expressions in pigs: first explorations of a novel methodology for integrative animal welfare measurement. Appl. Anim. Behav. Sci. 67 (3), 193–215. https://doi.org/10.1016/S0168-1591(99)00093-3.
Whitham, J.C., Wielebnowski, N., 2009. Animal-based welfare monitoring: using keeper ratings as an assessment tool. Zoo Biol. 545–560. https://doi.org/10.1002/zoo.20281.
Yamanashi, Y., Nogami, E., Teramoto, M., Morimura, N., Hirata, S., 2018. Adult-adult social play in captive chimpanzees: is it indicative of positive animal welfare? Appl. Anim. Behav. Sci. 199, 75–83. https://doi.org/10.1016/j.applanim.2017.10.006.
Yeates, J.W., Main, D.C.J., 2008. Assessment of positive welfare: a review. Vet. J. 175 (3), 293–300. https://doi.org/10.1016/j.tvjl.2007.05.009.